Webster J, Whitworth A, Morris J. (2015)

Is it time to stop 'fishing'? A review of generalisation following aphasia intervention.

Aphasiology

DOI: 10.1080/02687038.2015.1027169

Title: Is it time to stop 'fishing'? A review of generalisation following aphasia intervention

Authors: Webster, J.[1], Whitworth, A.[2] & Morris J.[1]

Addresses: [1] Speech and Language Sciences, School of Education, Communication and Language Sciences, King George VI Building, Newcastle University, Newcastle upon Tyne, NE1 7RU.

[2] School of Psychology and Speech Pathology, Curtin University, Western Australia

Short Title: Generalisation in aphasia intervention

Corresponding Author: Janet Webster

Contact Details: Speech and Language Sciences, School of Education, Communication and Language Sciences, King George VI Building, Newcastle University, Newcastle upon Tyne, NE1 7RU.

Telephone: 0191 222 5235 Fax: 0191 222 6518

Email: janet.webster@newcastle.ac.uk

Is it time to stop 'fishing'? A review of generalisation following aphasia intervention

Webster, J.[1], Whitworth, A.[2] & Morris J.[1]

[1] Speech and Language Sciences, Newcastle University, United Kingdom

[2] School of Psychology and Speech Pathology, Curtin University, Western Australia

Abstract

Background

The study of generalisation is integral to both demonstrating and maximising therapy effectiveness. With aphasia therapy aiming to promote change in everyday communication, there is a temptation to 'fish' for any evidence of improvement; multiple tasks, elicitation methods and measures may be used to identify any change which might then be attributed to therapy. Examining the evidence about generalisation will allow us to develop our theoretical understanding of change following aphasia intervention and strengthen our ability to form robust predictions for generalisation that can be tested.

Aims

We explore what is meant by generalisation, considering it within the broader context of impact of intervention. We propose a framework focusing on linguistic generalisation at the levels of word, sentence and connected speech. The discussion draws on the therapy literature for spoken production, exploring single word therapies for nouns and verbs, sentence production therapies and discourse therapies.

Main Contribution

The paper introduces a framework to conceptualise and describe generalisation *'within level'* i.e. change to untreated stimuli within the same linguistic level as the focus of treatment and *'across level'* i.e. change at a different linguistic level to the focus of treatment. The existing evidence base

for spoken production is reviewed using the framework, considering our current knowledge, the predictions we are able to make and areas for future focus.

Conclusions

There is an urgent need for further research in this area. While within level generalisation has often been addressed systematically, there have been more limited attempts to systematically define, predict and measure generalisation across linguistic levels. This has resulted in a limited evidence base from which to make decisions regarding when and how to best facilitate generalisation and where and how to best measure generalisation. We propose it is time to stop fishing for change in an opportunistic way and to develop a greater understanding of the relationship between change at the different linguistic levels of words, sentences and connected speech. We need to use theory and evidence to predict change (selecting the right 'pond' within which to fish) and identify valid and reliable ways to measure both targeted and generalised therapy effects (selecting the right 'bait').

Introduction

Generalisation is of central interest to clinicians and researchers in aphasia therapy as it is firmly embedded in the discussion of both demonstrating and maximising therapy effectiveness. We want to know that therapy is 'making a difference' to the individual with the benefits of our intervention extending, or generalising, beyond the focus of our therapy targets. Within spoken production, the generalisation of our treatment effects to everyday conversation is often viewed as the standard for determining whether our interventions have been successful (e.g. Carragher, Conroy, Sage, & Wilkinson, 2012; Lind, Kristoffersen, Moen, & Simonsen, 2009), irrespective of our understanding as to whether this is predicted. The temptation is, therefore, to 'fish' for any evidence of change in communication to demonstrate generalisation of therapy. A variety of constrained tasks, methods to elicit connected speech and analyses of those samples may be employed with the aim of identifying

changes which are then attributed to therapy. We propose it is time to stop fishing in this opportunistic way and to develop a greater understanding of generalisation by considering the relationship between changes at the different linguistic levels of words, sentences and connected speech. This will ensure a more purposeful and, ultimately, productive approach with theoretically driven predictions of change and a greater appreciation of how and where we measure generalisation of treatment outcomes.

This paper explores what is meant by generalisation and introduces a framework which can be used to conceptualise and describe linguistic generalisation, within a broader context of impact of intervention. We propose conceiving generalisation along a linguistic continuum, with generalised change possible at different linguistic levels of word, sentence and connected speech. Generalisation is considered as '*within level*' i.e. change at the same linguistic level as the focus of treatment and '*across level*' i.e. change at a different linguistic level to the focus of treatment. The paper applies the framework to existing studies of therapy for spoken production, considering treatments which focus at a single linguistic level (word, sentence or discourse) and multi-level treatments which target multiple linguistic levels, often with the aim of maximising treatment outcome. We consider the predictions we can make from the current evidence and those areas requiring further research, and explore issues that we believe need to be addressed if we are to understand the mechanism of the therapies we are employing, maximise therapy gains and demonstrate meaningful treatment effectiveness. We suggest that it is time to stop fishing for any change and instead use theory and evidence to predict the changes we should see, i.e. pay closer attention to selecting the right level or 'pond' into which to cast our line, and identify valid and reliable ways to measure the predicted therapy effects, including generalisation i.e. select the bait that is best placed to demonstrate whether or not intervention has been effective.

Defining generalisation

Within aphasiology, the term 'generalisation' lacks clear definition, with a range of terms used interchangeably. Generalisation can mean diverse things; at times, we compare performance on treated and untreated items, sometimes we employ tasks that are more closely related to everyday communication and, at other times, we seek to engage other people and other settings to 'facilitate' generalisation to other contexts. Within the wider literature, terms such as 'transfer' and 'carryover' are often used synonymously. Generalisation has also been used to refer to cross-modality generalisation, e.g. from comprehension to production, and cross linguistic generalisation (Kiran, Sandberg, Gray, Ascenso, & Kester, 2013). The potential is rife to muddy the waters through our lack of specificity in the use of this term and divert attention away from truly understanding the processes involved.

The behavioural literature, adopted early by aphasiologists, offers a contrast between 'response generalisation' and 'stimulus generalisation' (see Thompson, 1989 for a discussion and early review). Response generalisation is the emergence of an untrained language response following intervention while, in contrast, stimulus generalisation is the transfer of a trained behaviour to another context. Both might be a desirable outcome for the same or different treatments. For example, 'response generalisation' following word retrieval therapy might involve improvement both on treated words and untreated words. Stimulus generalisation might be limited to the specific items treated in therapy but the 'stimulus' may vary such that accurate retrieval of that response is evoked under different conditions; a lexical item may be retrieved, for example, in the context of an everyday scenario rather than in relation to the picture seen within the clinical setting. Thompson highlights the importance of generalisation to aphasia therapy, within the behavioural framework, suggesting:

"if response generalisation (the emergence of untrained language responses) does not occur as a result of treatment, then, in theory, clinicians must endeavour to train all responses that the aphasic patient will use. Further, if stimulus generalisation (the transfer of trained behaviors to stimulus conditions or situations that differ from those in which training takes place), does not occur,

treatment may be deemed unsatisfactory, since it is this carry-over of responding from the clinic to natural settings that is the ultimate goal of any rehabilitation programme" (Thompson, 1989, pp196).

While Thompson's points are still relevant today, we are now more aware of the complexities involved in both defining and unravelling generalisation as a process and considering its relationship to wider benefits of intervention.

Describing generalisation: A framework

We propose conceiving generalisation along a linguistic continuum, with *generalised change* seen at the different linguistic levels (i.e. word, sentence, connected speech). Given the diversity of areas, the continuum is not seen as one of complexity but of linguistic contexts. We differentiate this from the *impact of change*, evidenced by the person's or others' judgement or perception, intimately related but not a part of this continuum. Table 1 sets out a framework that considers the linguistic level, the elicitation methods that are employed and the area of focus. While connected speech has been described by others in different, but no less acceptable, ways, e.g. spontaneous versus semi-spontaneous speech (Prins & Bastiaanse, 2004), we have chosen to divide it according to the elicitation condition. Connected speech is elicited via (1) picture description, e.g. complex pictures or picture sequences, (2) monologues, such as production of a specific discourse genre, e.g. narrative, and (3) dialogues, such as seen in conversation. We propose that the latter two constitute a working definition of discourse, capturing the more complex, but different, demands of everyday speaking situations. In terms of focus, we have considered linguistic change in terms of lexical change, structural change (change in sentence structure or discourse structure), informativeness, i.e. the ability to convey a message (see Doyle, Goda, & Spencer, 1995; Nicholas & Brookshire, 1993) and changes in interaction (see table 1). *Within* level generalisation is defined as change to untreated stimuli within the same linguistic level as the focus of treatment e.g. generalisation from treated words to untreated words in a single word naming task, generalisation from treated sentence structures to untreated structures in a sentence production task. Within level generalisation will

normally be response generalisation as the elicitation stimulus is generally the same. *Across* level generalisation is defined as change at a linguistic level different to the focus of treatment e.g. production of words treated at single word level in sentences, production of treated/untreated syntactic structures in connected speech following sentence level therapy. The change across levels generally involves a change of stimulus/condition so it will be stimulus generalisation. Consideration of what we have defined as 'impact' is no less important but beyond the scope of this paper.

(insert table 1 here)

Exploring the Evidence

The above framework is used to review the current evidence regarding therapies for spoken production, considering within and across level generalisation and exploring single word therapies, sentence production and discourse therapies. Two distinctions are drawn here. The first is between a) single level treatment - treatments that target a single linguistic level and b) multi-level treatment – treatments that target more than one level. The second is between studies that measure within level generalisation only and those which measure both within and across level generalisation. Due to the large body of literature, the review is selective and should be read as an application of the framework rather than an exhaustive or systematic review of treatment approaches. It is, however, considered representative and therefore has been used to identify the predictions we are currently able to make about generalisation and areas for future focus. Therapy literature was searched according to the level of treatment (word, sentence and discourse) and for studies with a linguistic (rather than compensatory or pragmatic) focus. Where review papers were available, these were taken first, to consider if there was discussion of generalisation and whether themes were identified. Key therapy studies were also referred to when they investigated generalisation, particularly across level generalisation. The literature contained a large number of single case studies, some case series and some group studies; where therapy approaches overlapped, evidence from case series and groups studies was used as there had been replication across participants. Within the literature, a

number of aspects were considered in detail including (a) the focus of therapy, (b) the overall treatment approach i.e. whether it was strategic or targeted the underlying difficulty, and (c) the measures used to monitor outcome, including generalisation.

Single level treatment

This section will consider the outcome of therapies targeting a single linguistic level, describing word, sentence and discourse therapies in turn. Understanding within level change is important for understanding mechanisms of change and importantly for predicting across level generalisation.

Word level therapies

*Within level generalisation**:*** In therapy studies for noun and verb retrieval, there is perhaps the largest evidence of systematic study of within level generalisation i.e. whether improvement extends to untreated words. Scrutiny of findings, however, highlights variability in the extent of generalisation. In studies where strategic approaches have been used, generalisation is frequently reported. For example, participants have been taught to use access to intact orthographic information to support spoken production, including teaching grapheme to phoneme conversion to produce self-generated phonemic cues (e.g. Nickels, 1992), use of cueing aids (e.g. Bruce & Howard, 1987) or use of a letter board (e.g. Howard & Harding, 1998). Of the 19 participants described in Nickels' (2002) review paper who received strategic therapy, 13 showed generalisation to untreated items, emphasising that once a strategy was learned, it could be applied across items; the success of the strategy was mainly limited by co-occurring difficulties (Nickels, 1992). Where therapy has focused on directly improving impaired word retrieval (see table 2 for summary of case series studies), limited numbers of participants (21/69 in studies reported in table 2) show some generalisation to untreated words. On closer examination, however, these numbers may be reduced further, as in some cases, generalisation may reflect variability in performance across multiple testing periods as gains are not specifically related to the period of treatment (Fillingham, Sage, &

Lambon Ralph, 2006). Other studies (e.g. Carragher, Sage, & Conroy, 2013) highlight that gains on untreated items are not maintained over time.

The underlying nature of the impairment is one factor that may influence whether generalisation is seen. Best and colleagues (Best et al., 2013) analysed the data from two studies investigating the effect of phonological cues and orthographic cues (described in Best, Greenwood, Grassly, & Hickin, 2008; Hickin, Best, Herbert, Howard, & Osborne, 2002) to determine whether generalisation to untreated items was related to participants' language profiles. Their analysis showed there was minimal change in the naming of untreated items across all participants but generalisation was more likely to be seen in participants with relatively good semantics and poor phonological encoding, i.e. a post-lexical deficit identified by a high proportion of phonological errors and the presence of a length effect. This profile is consistent with other participants (e.g. Fisher, Wilshire, & Ponsford, 2009) who show within level generalisation. Other participants, for example, in the Leonard et al. (2008) and Raymer et al. (2012) studies, and participant GF (Robson, Marshall, Pring, & Chiat, 1998) also produced phonological errors and showed generalisation, although the authors attributed their difficulties to a lexical deficit. In these studies, generalised gains were attributed to improved lexical access (Robson et al., 1998) or spreading activation within the phonological system (Leonard et al., 2008). It should also be noted that, in a study of four participants with a combination of lexical and post-lexical phonological impairments, generalised improvement was not seen (Waldron, Whitworth, & Howard, 2011).

(insert table 2)

There has also been considerable debate as to whether the focus of therapy and the therapy task influences generalisation. Initial case studies of semantic-based therapy involving selection or generation of semantic features reported generalisation to untreated words (see Boyle (2010) for a detailed review of these studies). A notable feature of these studies, however, is that they all involved small word sets where the gains in word retrieval arose from small changes in the number

of words retrieved. They also involved repeated probing and therefore gains on untreated words may have been a consequence of repeated attempts to name rather than generalisation per se (Boyle, 2010; Howard, 2000; Nickels, 2002).

Across level generalisation: Across level generalisation (within single level treatments at word level) refers to when change is seen in the use of treated and/or untreated words at sentence level or within connected speech. Despite the large number of studies investigating therapy for noun retrieval, there has been minimal consideration of across level generalisation, with only isolated examples in single case studies. Rose and Douglas (2008) investigated generalisation across contexts using procedural discourse designed to elicit words within treated categories, e.g. describing 'going to the zoo' to elicit 'animals'. Single word gains in the retrieval of treated words were associated with an increased number of nouns in discourse. A number of studies have used Nicholas and Brookshire's correct information unit (CIU) analysis (Nicholas & Brookshire, 1993) to investigate changes in the informativeness of connected speech (picture description and monologues) following word retrieval therapy. Boyle and Coelho (1995) investigated changes in connected speech following SFA therapy; no across level generalisation was seen, with no change in mean words per minute, mean CIU per minute or percentage words that were correct CIU. Another study (Spencer et al., 2000) did find an increase in percentage of CIUs after phonological therapy resulting in generalised gains at single word level. Best, Howard, Bruce and Gatehouse (1997) reported gains in communicative effectiveness (as rated by naïve listeners) following their strategic cue therapy and there was also an increase in the number of content words produced during picture description; these gains were seen alongside generalised gains across treated and untreated items at single word level. Generalised across level effects of therapy for phonological difficulties have also been reported by Franklin, Buerk and Howard (2002) and Rose, Douglas and Matyas (2002), with both studies reporting a reduction in phonological errors in connected speech (narrative or conversation respectively). In the latter study, content words per substantive turn was also measured and did not increase.

There has been greater consideration of across level generalisation in studies of verb retrieval, possibly reflecting the integral role that verbs play in sentence production. Studies have investigated gains at sentence level and/or in connected speech. Within level gains in the retrieval of treated verbs have been accompanied by improvement in sentence production around treated verbs (e.g. Marshall, Pring, & Chiat, 1998; Raymer & Ellsworth, 2002). More general gains are also seen around untreated verbs (Marshall et al., 1998) or on general tests of sentence production (Boo & Rose, 2011; Rose & Douglas, 2008). Gains appear to be independent of the nature of therapy and the underlying verb deficit although co-occurring syntactic deficits can restrict improvement (e.g. Mitchum & Berndt, 1994; Raymer & Ellsworth, 2002). In conversation, the participants in Carragher et al.'s study (2013) showed no significant change in the number of verbs retrieved following verb therapy. There was no correlation between verb naming ability and verb retrieval in conversation or between improvement in verb naming for either treated or untreated verbs and change in the number of verbs retrieved in conversation. There have, however, been some individual participants who have shown an increase in the number of verbs produced in picture description (Boo & Rose, 2011; Rose & Sussmilch, 2008) and in conversation (Carragher et al., 2013; Rose & Sussmilch, 2008).

Key Themes: Current evidence suggests that within level generalisation at the single word level is influenced by treatment approach, the focus of therapy and the underlying impairment. No difference is seen between studies targeting verb or noun retrieval. If treatment teaches a strategy, widespread generalisation to untreated words is more likely to be seen (Nickels, 2002). If the focus is on strengthening the link between word meaning and word form, no generalisation is predicted (Howard, 2000); this item specific change is seen in the majority of participants. For people with phonological deficits, particularly those with post-lexical difficulties, generalisation to untreated words is sometimes seen (Best et al., 2013). There is currently minimal evidence to suggest differential gains as a consequence of specific therapy techniques although some tasks may result in participants developing their own strategies. If there is an assumption that there is spreading activation between words that share either semantic or phonological features and tasks strengthen

those connections (Boyle & Coelho, 1995; Leonard et al., 2008), generalisation might be predicted to untreated words that share those features but with no generalisation to unrelated words; the evidence to support this, however, is currently limited by the design of the studies.

The paucity of investigations looking at the impact of noun retrieval therapy on the production of sentences and connected speech highlights the need for further research. With the majority of participants showing minimal within level generalisation to untreated nouns, it is predicted that across level gains will be restricted to treated nouns; this makes measurement very challenging with specific contexts needed to elicit the words (as in Rose & Sussmilch, 2008). In cases where within level generalisation is reported, wider change in lexical context or informativeness is predicted and has been seen in some participants; further investigation is, however, needed. Across level generalisation is more widespread following verb therapy, with wider changes in structural and lexical content; possible explanations for these changes will be considered in the later section exploring the effects of multi-level therapy on verb and sentence production.

Sentence level therapies

This section will consider generalisation as a consequence of sentence level treatment, focusing on two frequently reported therapy types:-verb-centred mapping therapies (see Marshall, 1995 for review) and Treatment of Underlying Forms (TUF) (see Thompson & Shapiro, 2007; Thompson, Shapiro, Kiran, & Sobecks, 2003, for detailed reviews). Generalisation has been considered within level, investigating improvement to different sentence types, and across level to more widespread gains in connected speech. The issue of generalisation for sentence level therapies is complicated by the fact that therapies often have a comprehension focus with the aim of improving sentence production; this constitutes cross-modality generalisation.

Within level generalisation: Verb-centred mapping therapies focus on identifying the verb and other sentence components in written sentences, with the aim of increasing awareness of sentence structure and thereby improving production. TUF, reported in a large number of studies by Thompson and colleagues, is an adapted version of mapping therapy for complex structures e.g. passives, questions, object relatives (see Thompson & Shapiro, 2007; Thompson et al., 2003). TUF involves the identification of sentence components in an active sentence, followed by instructions about how to manipulate the constituents to produce the complex form. Within level, some generalisation has been seen across sentence types following both therapies. For mapping therapy, the most prominent form of generalisation reported is from verb based transitive structures (e.g. the doctor shoots the vicar) to reversible locative structures (e.g. the cloud is above the plane) and vice versa (Byng & Coltheart, 1986; Haendiges, Berndt, & Mitchum, 1996; Jones, 1986). TUF studies are based on strong theoretical predictions that, following treatment, gains will be seen across structures that are linguistically similar, with no gains in linguistically different structures; studies have shown this to be evident. For example, studies have reported that training structures involving 'wh' movement (e.g. object 'wh' questions, object cleft) results in gains to other structures with 'wh' movement but no change in structures involving noun phrase movement (e.g. passives, subject raising sentences) and vice versa. Generalisation was enhanced when treatment started with the most complex structures (Thompson et al., 2003). This led Thompson to propose the Complexity Account of Treatment Efficacy (CATE) which states that training complex structures results in generalisation to less complex structures when/if untrained structures encompass processes relevant to the treated ones. No generalisation is seen from simple to complex structures.

Across level generalisation: In therapies that have focused on sentence production, across level gains have been reported in connected speech, primarily narratives. At the broadest level for the verb centred mapping therapies and TUF, this shows cross-modality generalisation, with therapy focused on comprehension resulting in gains in production. The mapping therapies have resulted in an increased use of verbs and in an increase in verb argument structures, (Byng & Coltheart, 1986;

Jones, 1986) although differences are seen across participants (Byng, Nickels, & Black, 1994). Gains in connected speech are also reported following TUF (e.g. Ballard & Thompson, 1999), with improvement on both lexical and structural measures for some participants; quantitative gains were accompanied by improved ratings from naïve listeners on measures of content, coherence and fluency.

Key Themes: Within level generalisation has been demonstrated following sentence level therapy, although this is dependent on the linguistic similarity between and relative complexity of treated and untreated structures. Across level generalisation to connected speech has also been seen and may reflect either use of the sentences which improve at sentence level or wider gains across sentence types. There is not necessarily an increase in the use of targeted structures following mapping therapy or TUF (Mitchum, Greenwald, & Berndt, 2000) but these structures e.g. subject relatives, object clefts etc. are low frequency structures and may not be elicited within the samples (Roland, Dick, & Elman, 2007). Alternatively, the metalinguistic focus may increase awareness of the sentence components and provide a strategy for producing them, resulting in wider gains across sentences; if this is the case, the generalisation may reflect the use of the strategy rather than improved thematic mapping (Mitchum et al., 2000).

Discourse level therapies

When considering generalisation of intervention, the discussion would not be complete without exploring the discourse level. While considerable attention has been directed to how narrative structure can be analysed and whether discourse elements are impaired in aphasia (e.g. Glosser & Deser, 1991; Ulatowska, North, & Macaluso-Haynes, 1981), a limited number of studies have attempted to intervene in these behaviours and/or subsequently measured whether generalisation takes place to untreated discourse. Within level, we need to explore whether discourse therapies generalise to other discourse tasks or to contexts that are not specifically treated.

Within level generalisation: Some examples are present in the literature where whole task training has been used and within level generalisation has been measured. Hinckley and Carr (2005), for example, focused on improving communication in a personally relevant discourse task (ordering from a catalogue). Therapy involved repeated practice of the trained discourse task, with the person problem solving and developing strategies to achieve effective communication. Following therapy, significant improvement was seen in the trained discourse task but there was no generalisation to performance on the CADL-2 (Holland, Frattali, & Fromm, 1999) which involves other everyday scenarios; generalisation to related discourse tasks was therefore not seen. Within level generalisation was also explored by Youmans, Holland, Muñoz and Bourgeois (2005) who evaluated the effect of script training for personally relevant discourse. Training moved from repeated repetition of phrases to independent production of script. Following therapy, trained scripts became more automatic, more fluently produced and natural, with increased flexibility (i.e. some variation from initial script). Generalisation was seen with some use of the trained script in conversation and with novel conversation partners. There was, however, no generalisation to untrained discourse topics.

Key Themes: There has been limited investigation of within level generalisation at discourse level. Where it has been considered, there has been no generalisation to untrained discourse tasks or topics. One possible explanation contributing to this may be the relatively underdeveloped frameworks available for measuring different discourse, along with the limited availability of assessment tools at the discourse level that permit therapy targets to be measured in novel (within level) contexts or across genres.

Multi-level treatment

This section will consider generalisation as a consequence of treatments which work across multiple linguistic levels. The treatments aim to maximise change by providing a more natural context for

therapy and/or promoting generalisation to everyday communication. The section is divided into different types of studies:- i) word and sentence ii) word and connected speech iii) sentence and discourse and iv) word, sentence and discourse. As therapy occurs across levels in these studies, it is often difficult to determine whether generalisation is within or across level, particularly if treatments occur simultaneously.

Word and sentence level therapies

There are a number of therapies which combine work on verb and sentence production, reflecting the centrality of verbs for sentence production. Bastiaanse and colleagues (Bastiaanse, Hurkmans, & Links, 2006; Links, Hurkmans, & Bastiaanse, 2010) have suggested that verbs should not be treated in isolation and that sentences should always be the starting context for therapy. Combined approaches have also been selected due to the co-occurrence of verb and argument structure difficulties (Webster, Morris, & Franklin, 2005). Therapy techniques are diverse, including treating verbs within a sentence context, combining work on verb retrieval and sentence production or aiming to increase awareness of verb argument structure by focusing on verbs and their associated nouns. A detailed description of the varied therapies is beyond the scope of this paper (see Conroy, Sage, & Lambon Ralph, 2006; Webster & Whitworth, 2012 for reviews).

Webster and Whitworth (2012) reviewed individual therapy studies and reported improvement at single word level, in sentence production and in connected speech. This review demonstrates that, as with studies of single word verb retrieval therapy, the studies with a combined focus on verbs and sentences result in item specific gains at word level. At sentence level, some studies/participants show structural improvements in sentence production, with significant gains on general sentence production tests (e.g. Edwards & Tucker, 2006; Links et al., 2010) and in sentence production around both treated and untreated verbs (e.g. Edmonds, Nadeau, & Kiran, 2009; Schneider & Thompson, 2003). Across level, gains in connected speech are reported with change in lexical content (e.g. Links

et al., 2010), on structural measures (e.g. Edmonds et al., 2009; Links et al., 2010; Schneider & Thompson, 2003) and informativeness (e.g. Wambaugh & Ferguson, 2007). There is, however, extensive variability in the elicitation methods, the specific measures used to capture lexical and structural change and the extent to which change is seen across participants, across elicitation methods and across measures.

Key Themes: There are similarities between the outcome of single word verb therapy and multi-level therapy combining verb and sentence therapy, and hence, this discussion considers both. Specific gains at single word level are accompanied by gains in the use of treated verbs at sentence level; this provides robust evidence that improving lexical access at single word level also improves lexical access at sentence level. As with nouns, given the minimal generalisation to untreated verbs in a single word context, it might be predicted that across-level gains would be restricted to the retrieval of treated verbs (Conroy, Sage, & Lambon Ralph, 2009c). However, as previously highlighted there seems to be a difference between nouns and verbs. Following verb therapy (whether single level or multi-level), some generalised lexical gains at sentence level and in connected speech are reported and these are accompanied by gains on structural measures i.e. improved sentence production. This suggests the predictions for across level generalisation for nouns and verbs are different; this could reflect the centrality of verbs for sentence production but the mechanism for generalisation remains unclear. Edwards and Tucker (2006) suggested that general improvements in sentence production should be seen as therapy is enhancing the retrieval of the category of verbs rather than specific verbs. Improving lexical access may benefit sentence production by freeing up processing resources (Linebarger, McCall, & Berndt, 2004); this would predict generalised gains although there may still be a difference between sentences containing treated and untreated verbs (as in Schneider & Thompson, 2003). Alternatively, combining work on verbs and nouns may engage argument structure information, thereby maximising potential generalisation to sentence production (Mitchum et al., 2000). In this case, gains may extend beyond treated verbs and nouns if therapy increases the participant's awareness of the role of verbs in sentences and provides a general

strategy to specify the argument structure (thinking about 'who', 'what to', 'what with' and 'where') (Webster et al., 2005). Although there is evidence of across level generalisation on both lexical and structural measures, the different elicitation methods and measures used make it difficult to compare across studies; there is a need, therefore, to systematically compare the outcome of single word therapy and the different types of multi-level therapy.

Word and connected speech level therapies

There have been a number of studies which have worked on word retrieval with the explicit involvement of connected speech as one of the contexts for therapy (see Boyle, 2011 for a review). Herbert and colleagues (Herbert, Best, Hickin, Howard, & Osborne, 2003) reported on therapy combining traditional noun therapy involving orthographic and phonological cues with 'interactional' therapy. Interactional therapy was designed to elicit target words in a more naturalistic context, e.g. producing a shopping list or in conversation, potentially facilitating generalisation across levels to everyday communication. The effect of interactional therapy was considered in relation to gains in the naming of treated and untreated items as well as in a task assessing production of targeted nouns in more everyday scenarios. Therapy resulted in item specific gains in picture naming for five of the six participants and an increase in the communicative use of the target words. However, contrary to Herbert et al. (2003), minimal gains were seen following interactional therapy in a subsequent case series study involving eight participants (Best et al., 2008).

Several studies have combined semantic feature analysis (SFA) and discourse tasks (Antonucci, 2009; Boyle, 2004; Peach & Reuter, 2010). These studies used primarily picture based tasks, either describing sequences (Boyle, 2004; Peach & Reuter, 2010) or within a barrier task in a group (Antonucci, 2009). SFA was used when word retrieval difficulties were evident, either as they arose within the context of the discourse (Antonucci, 2009; Boyle, 2004) or following the discourse task (Peach & Reuter, 2010). The studies, therefore, varied in the extent particular words were treated

using SFA. The treated words did not necessarily show improvement in a single word context but gains in connected speech were seen, with improvement beyond the specific words treated. In each of the studies, participants showed change in either the number or percentage of CIU (Nicholas & Brookshire, 1993), resulting in greater informativeness.

Key Themes: The reported variations in therapy outcome highlights the need for further investigation into the relationship between lexical retrieval at single word level and in connected speech. The use of more natural everyday contexts may be needed to promote generalisation of the use of targeted words in connected speech but may not be sufficient (without structured repetitive tasks) to produce specific gains in word retrieval. Overall gains in informativeness are seen in the studies combining SFA and discourse tasks. While Boyle (2011) suggests this is 'reflective of a generalised improvement in word retrieval abilities' (p. 1322), the mechanism for the improvement and the extent to which the discourse component was crucial for producing the gains is not clear.

Sentence and discourse level therapies

There is limited evidence currently describing therapy involving the sentence and discourse level. Murray, Timberlake and Eberle (2007), in their single case study, introduced a 'discourse training module' to their TUF therapy protocol, with the aim of enhancing generalisation to discourse and increasing the communicative value of therapy. Targeted syntactic structures were embedded within a discourse framework, with stimuli from newspapers and magazines. Therapy focused on training written sentence and discourse production but as the effects were measured in both spoken and written discourse (narrative and procedural discourse), the study is included here. Therapy resulted in the same patterns of within level generalisation as traditional TUF therapy; gains were seen in the written production of treated structures and linguistically similar untreated structures with no improvement in unrelated structures. Similar cross-modality gains were seen in spoken production. At discourse level, gains were also seen in both spoken and written discourse, with lexical (number of substantive verbs) and structural (MLU) changes and improved informativeness (number and

percentage of CIU). As changes in connected speech are seen following traditional TUF, it is not clear whether the discourse component was necessary to produce these gains.

Word, sentence and discourse level therapies

We now describe studies that combine work on words, sentences and discourse, either within a single task or across tasks used concurrently. Milman, Vega-Mendoza and Clendenen (2014) used a multi-level approach, Integrated Training for Aphasia (ITA), with three participants with aphasia. ITA involved the delivery of word, sentence and discourse level therapy (20 minutes each of a one hour session) where the single vocabulary items were incorporated into the sentence and discourse components, and the morphosyntactic structures included in discourse. The targets, however, were words and sentences rather than any direct focus on discourse features. Within level generalisation was not found at the word level for semantically related words and, across level, the relationship between change at word and sentence levels was not clear. Perhaps of most interest here, however, was that all participants demonstrated significant gains on a variety of lexical and structural measures in picture description (e.g. one or more of MLU, noun/verb ratio and open/closed class word ratio). Whitworth, et al. (this volume) report on a multi-level approach, NARNIA (Novel Approach to Real-life communication: Narrative Intervention in Aphasia) that combines word retrieval (with a focus on verbs and then associated nouns), sentence production (generic argument structure) and discourse macrostructure of everyday speaking situations within a single task. Findings from this study showed that within level generalisation was seen at both the word and discourse levels, with gains seen in naming of untreated verbs and in the use of taught discourse structures across a range of everyday discourse genres. No within level generalisation was seen at the sentence level. With respect to across level generalisation, gains in single word (verb use) and sentence production were evident in discourse. What is noteworthy about this study was the systematic attempt to measure both within and across level generalisation at each of the language levels, including discourse.

Key Themes: These multi-level therapies have shown gains on connected speech measures, emphasising their potential to influence communication. The fact that these therapies are-multi-component can make it difficult to tease out what aspects of the intervention, and in what combination, are influential to the outcome.

Exploring the Issues

The review of the above evidence demonstrates that there are clear directions offered in the literature to date, particularly with respect to within level generalisation, and quite systematic attempts being made to progress our understanding of this complex area. It is also clear, however, that the evidence base remains mixed, with particular questions remaining around prediction and measurement.

Using theory and evidence to predict change: selecting the right pond

If we return to the fishing analogy, when thinking about the context for measuring generalisation we need to consider into which pond to cast our line. We need to approach this from a theoretically motivated, evidence based position, both predicting in advance what generalisation of treatment effects we expect and having clear hypotheses why. In many instances, it *is* highly appropriate to look within level for evidence that therapy has worked - this will extend and build our understanding of the connectivity between words, the relationship between sentence types and different discourse genres. We may also seek to examine the effect of therapy across linguistic levels, being both led by the evidence and by our linguistic theories. When therapies are multi-level, the distinction between within and across level generalisation may be blurred, particularly if levels are targeted simultaneously rather than sequentially.

With respect to within level effects, our review has highlighted that generalisation has been reported as a consequence of single word, sentence and discourse therapies but that it is not

universal. The importance of this, as Thompson highlighted in relation to response generalisation, is that if within level generalisation does not occur, we need to train all of the responses (i.e. all of the words or sentence structures needed). As this is often an unrealistic aim, we need to, crucially, focus on what is functionally useful. At the single word level, there is more evidence of generalisation when therapy involves the use of a strategy and, if a strategic therapy is appropriate (based on evidence and pre-requisite abilities), this may maximise therapy gains. With lexical therapies which result in items specific gains, it is crucial to choose materials that are relevant to the person (see Renvall, Nickels, & Davidson, 2013a; 2013b for detailed discussion). Appreciating this distinction does not detract from lexical therapies targeting specific words, especially as people with aphasia frequently have a need to learn particular lexical items, and there is robust evidence that lexical access can be improved with practice. The capacity of individuals to learn strategies, often related to their residual language functions, is also variable, a factor that will heavily influence the therapy approach. At the sentence level, evidence suggests we need a robust understanding of the linguistic similarity between structures, possibly focusing on the most complex structures first (Thompson & Shapiro, 2007). Discourse remains a relatively untapped area.

From the current evidence, across level generalisation is more difficult to predict but it is an awareness and understanding of these changes that is a crucial aspect of evaluating the communicative value of therapy. From single level therapy, our review has shown that there has been limited investigation of the generalisation of noun retrieval therapy to sentence production and connected speech. There is some evidence of treated words being used more frequently but measuring these specific gains has been, and remains, challenging. There is also limited evidence of more generalised gains in conversation but these are only present where participants have shown generalised gains across treated and untreated nouns at single word level. Single word verb therapies have been shown to result in improved sentence production but, again, change in connected speech has received fairly minimal attention. Where connected speech has been

measured following sentence level therapies, both lexical and structural gains have been reported. Across the studies, the importance of considering an individual's communication profile, their strengths and co-occurring difficulties has been emphasised and is likely to contribute to variability in therapy outcome; this reinforces the value of studies where there are multiple participants profiled and studied.

Gains across levels have been considered more comprehensively in studies investigating multi-level therapies. These studies are, however, very diverse and highlight the complexity of studying therapy effects and generalisation. Some therapies work primarily at one level with the subsequent introduction of tasks to promote generalisation of single word gains to everyday communication. Some therapy tasks work across levels (within a single task), for example, combining work on verbs and sentences or embedding work on word retrieval and sentence production within a discourse context. Some studies describe concurrent use of different tasks which work across level. Milman et al., (2014) advocate three key reasons for using therapies that incorporate different linguistic levels- the interconnected nature of linguistic networks, the potential facilitation of generalisation to everyday communication and the reality of clients having multiple difficulties. Interventions spanning different levels potentially allow multiple difficulties to be tackled simultaneously and recognise the relationship between levels. However, understanding the benefits of multi-level interventions relies on our ability to describe the therapy components, identify therapy gains, and where possible, directly link the change(s) to a component of the therapy.

Measuring predicted generalisation: selecting the right bait

When we have identified the context (the pond), we still need to consider how to measure therapy effects, including generalisation, i.e. which bait to use. In measuring change, it can be tempting to seek evidence of generalisation using a wide range of measures with the hope that something will

show change. This non-discriminatory sampling may identify change due to the amount of comparisons and measures but will not further our understanding as to why change took place. This is quite different to starting therapy with clear predictions about the change expected (within and across levels) and using reliable and valid measures to assess relevant behaviours and predicted change (within and across levels).

When assessing within-level generalisation, a range of constrained tasks have been used. There is generally a high degree of similarity between the therapy task and assessment task although the stimuli may vary in complexity, e.g. typical/atypical words, simple/complex sentences. As tasks are designed to identify specific change in treated words/structures and specific patterns of within level generalisation, standardised assessments are generally not used. Constrained tasks, such as picture naming are easy to administer, can be well controlled and have a high degree of test re-test reliability (Herbert, Hickin, Howard, Osborne, & Best, 2008). There is, however, debate as to whether the constrained tasks are ecologically valid (Ferguson & Armstrong, 1996; Herbert et al., 2008; Mayer & Murray, 2003). In order to address this, some studies (Herbert et al., 2008; Mayer & Murray, 2003) have looked at the relationship between lexical retrieval across contexts, with contrasting findings and conclusions. Herbert et al. (2008) showed a strong correlation between single word naming scores and lexical retrieval measures in conversation. They concluded 'that word retrieval in picture description and conversation are quantitatively related' (p195) and that picture naming was a valid assessment of word retrieval. In contrast, Mayer and Murray (2003) found performance on a naming task was a predictor of aphasia severity but not of word retrieval in picture description and conversation. While these studies do consider the extent to which processes are shared across levels, it remains unclear as to what this correlational data can add to the debate about generalisation of therapy effects. There has been limited investigation of the relationship between other measures e.g. constrained measures of sentence production and sentence production in connected speech, although there is evidence that the presentation of syntactic

difficulties may change across language testing and conversational contexts (Beeke, Wilkinson, & Maxim, 2003).

Measuring across level generalisation is more complex. There are inherent differences across language levels such that the change of task/stimulus, the demands on linguistic and cognitive processing (Conroy et al., 2009c) and the impact of co-occurring difficulties, all need to be considered. Conroy et al. (2009c) investigated whether gains following naming therapy for nouns and verbs were seen in picture naming, picture supported narratives and unsupported narratives. Post-therapy, they showed a step wise decrement in naming accuracy across the three contexts, arguing that the processes in lexical retrieval overlapped but that the production of narratives was more linguistically, cognitively and pragmatically more demanding. This may suggest a continuum with the relative complexity of the task influencing the extent to which change is seen.

Conversation as the gold standard: Is it the prize catch?

As discussed at the outset, in seeking to demonstrate generalisation, clinicians are encouraged to look for change in everyday communication, with conversation frequently viewed as the 'gold standard'. Analysing linguistic change in dialogic discourse in a reliable way is, however, difficult. We know, for example, that conversation depends on the nature and interaction of the topic, levels of interest and motivation, and the relationship between conversation partners, each influencing the linguistic content and resulting in variability across speakers and time (Carragher et al., 2012).There have been some attempts to measure lexical and structural changes in conversation following therapy (see Carragher et al., 2012 for review), with studies acknowledging the challenges posed by sample variability. However, many studies have used other elicitation contexts e.g. picture description or monologic discourse tasks e.g. story narrative, recount, procedure and exposition (see discourse protocol in Whitworth, Claessen, Leitão, & Webster, under review). With the increased constraint, there is more control over the sample (in terms of specific words, sentence and discourse

structure) and more reliability across different samples. While not claiming to capture the interactive nature of communication, these sampling contexts enable identification of features that are different to the production of healthy speakers, allow comparison between individuals, and promote monitoring of performance over time (Carragher et al., 2012; Lind et al., 2009). There is still, however, even within these contexts, considerable variation across speakers (including healthy speakers) (Webster, Franklin, & Howard, 2007) and across time, particularly in people with aphasia (Cameron, Wambaugh, & Mauszycki, 2010), such that test re-test reliability and sensitivity to both impairment and change may be limited. This reinforces the need for repeated sampling and emphasises that understanding variation is important in demonstrating gains as a consequence of generalisation (Carragher et al., 2012). The relative length of such prescriptive monologues, however, often being quite short, does introduce concern as to how representative they are of a person's ability (see discussion in Armstrong, 2000).

There is a need to understand the relationship between elicitation conditions in order to identify reliable and valid measures of improvement. Crucial differences exist between picture description, narrative and conversation which may influence the overall complexity of the task (Conroy et al., 2009c) as well as making specific demands on other aspects of processing (e.g. vision, memory). These differences include the picture support available (with the corresponding access to semantics), flexibility in terms of choice of words and/or sentence type, the contrasting demands of monologue and the dyadic nature of conversation (both in relation to the levels of structure required and the support, or otherwise, offered by the conversation partner), and the demands of comprehension. These different demands (both cognitive and interactional) mean that change may be differentially seen in conversation and other types of connected speech.

The lack of consensus around contexts and standardisation across elicitation methods is accompanied by uncertainty about which measures to use, i.e. what are the relevant variables that are sensitive to the predicted change and can they be measured in a reliable way, keeping in mind that different variables are likely to be associated with different linguistic change. Within the studies

reported earlier, different measures of lexical and structural gains have been shown to be differentially sensitive. For example, type token ratio may be capturing lexical diversity whereas proportion of words used only once may reflect lexical richness (Lind et al., 2009). The sensitivity of the measure chosen may also depend on the denominator used. For example, Herbert et al. (2008) showed a significant relationship between single word naming and content words as a proportion of turns or substantive turns but no relationship between single word naming and content words as a proportion of speech units. These examples highlight the complexity involved in identifying which measures are the most relevant to capture generalisation. Identifying measures that are relevant and simple and applicable to the clinical setting is also crucial (Lind et al., 2009). We need to understand the relationship between tasks, the measures that are sensitive to impairment and the relationship between specific lexical and structural measures and measures of informativeness and interaction. This will then enable us to understand the relationship between change across tasks and between linguistic levels. It may also enable us to reduce the amount of sampling we carry out, undertaking theoretically motivated assessment and yet saving considerable time and effort for the clinician, researcher and, importantly, person with aphasia.

Conclusions

The aim of therapy is to maximise gains in everyday communication, reduce the disability associated with aphasia, and promote increased participation. Within clinical services, therapists are accountable for demonstrating that therapy has been effective and has produced gains that are meaningful to the person with aphasia. This requires demonstration of both the specific linguistic gains that result from targeted therapy and the overall impact of therapy. This paper has focused on the linguistic gains seen within and across levels that demonstrate change beyond explicit therapy targets. We are currently limited in the extent to which we can predict the degree of generalisation that follows treatment; this is a direct consequence of our underdeveloped understanding of the relationship between linguistic levels and between change across those levels. A stronger

understanding will emerge from (i) applying knowledge of linguistic therapy to predictions (ii) systematic testing of predictions about across level generalisation (iii) intervention studies that investigate the effects of single level and multi-level therapies, with systematic consideration of within level and across level effects and, finally, (iv) our continued investigation of the relationship between constrained tasks and connected speech, and between different types of connected speech. The impact of therapy on the person's perception of their ability and overall wellbeing, and how this relates to any linguistic change, is no less important. Just as there is a need to understand generalisation across linguistic levels, the relationship between linguistic change and impact also needs systematic evaluation. A greater understanding of the relationship between tasks, between levels and between linguistic change and impact, should also allow us to measure therapy effectiveness in a more efficient way. It is time to stop 'fishing' in an indiscriminate way. The evidence base is building and our challenge is to make clear, well defined and specific predictions about generalisation and evaluate these with appropriate measurement tools.

References

Abel, S., Schultz, A., Radermacher, I., Willmes, K., & Huber, W. (2005). Decreasing and increasing cues in naming therapy for aphasia. *Aphasiology, 19*(9), 831-848.

Antonucci, S. M. (2009). Use of semantic feature analysis in group aphasia treatment. *Aphasiology, 23*(7-8), 854-866.

Armstrong, E. (2000). Aphasic discourse analysis: The story so far. *Aphasiology, 14*(9), 875-892.

Ballard, K. J., & Thompson, C. K. (1999). Treatment and generalization of complex sentence production in agrammatism. *Journal of Speech, Language, and Hearing Research, 42*(3), 690-707.

Bastiaanse, R., Hurkmans, J., & Links, P. (2006). The training of verb production in Broca's aphasia: A multiple-baseline across-behaviours study. *Aphasiology, 20*(02-04), 298-311.

Beeke, S., Wilkinson, R., & Maxim, J. (2003). Exploring aphasic grammar 2: do language testing and conversation tell a similar story? *Clinical linguistics & phonetics, 17*(2), 109-134.

Best, W., Greenwood, A., Grassly, J., Herbert, R., Hickin, J., & Howard, D. (2013). Aphasia rehabilitation: Does generalisation from anomia therapy occur and is it predictable? A case series study. *Cortex, 49*(9), 2345-2357. doi: 10.1016/j.cortex.2013.01.005

Best, W., Greenwood, A., Grassly, J., & Hickin, J. (2008). Bridging the gap: can impairment-based therapy for anomia have an impact at the psycho-social level? *International Journal of Language & Communication Disorders, 43*(4), 390-407.

Best, W., Howard, D., Bruce, C., & Gatehouse, C. (1997). Cueing the words: A single case study of treatments for anemia. *Neuropsychological Rehabilitation, 7*(2), 105-141.

Boo, M., & Rose, M. L. (2011). The efficacy of repetition, semantic, and gesture treatments for verb retrieval and use in Broca's aphasia. *Aphasiology, 25*(2), 154-175.

Boyle, M. (2004). *Discourse treatment for word retrieval impairment in chronic aphasia*. Paper presented at the 34th Clinical Aphasiology Conference, Park City, UT.

Boyle, M. (2010). Semantic feature analysis treatment for aphasic word retrieval impairments: What's in a name? *Topics in stroke rehabilitation, 17*(6), 411-422.

Boyle, M. (2011). Discourse treatment for word retrieval impairment in aphasia: The story so far. *Aphasiology, 25*(11), 1308-1326.

Boyle, M., & Coelho, C. A. (1995). Application of semantic feature analysis as a treatment for aphasic dysnomia. *American Journal of Speech-Language Pathology, 4*(4), 94-98.

Bruce, C., & Howard, D. (1987). Computer-generated phonemic cues: An effective aid for naming in aphasia. *International Journal of Language & Communication Disorders, 22*(3), 191-201.

Byng, S., & Coltheart, M. (1986). Aphasia therapy research: Methodological requirements and illustrative results. *Advances in Psychology, 34*, 191-213.

Byng, S., Nickels, L., & Black, M. (1994). Replicating therapy for mapping deficits in agrammatism: Remapping the deficit? *Aphasiology, 8*(4), 315-341.

Cameron, R. M., Wambaugh, J. L., & Mauszycki, S. C. (2010). Individual variability on discourse measures over repeated sampling times in persons with aphasia. *Aphasiology, 24*(6-8), 671-684.

Carragher, M., Conroy, P., Sage, K., & Wilkinson, R. (2012). Can impairment-focused therapy change the everyday conversations of people with aphasia? A review of the literature and future directions. *Aphasiology, 26*(7), 895-916.

Carragher, M., Sage, K., & Conroy, P. (2013). The effects of verb retrieval therapy for people with non-fluent aphasia: Evidence from assessment tasks and conversation. *Neuropsychological Rehabilitation, 23*(6), 846-887.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2006). Towards theory-driven therapies for aphasic verb impairments: A review of current theory and practice. *Aphasiology, 20*(12), 1159-1185.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009a). Errorless and errorful therapy for verb and noun naming in aphasia. *Aphasiology, 23*(11), 1311-1337.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009b). The effects of decreasing and increasing cue therapy on improving naming speed and accuracy for verbs and nouns in aphasia. *Aphasiology, 23*(6), 707-730.

Conroy, P., Sage, K., & Lambon Ralph, M. A. (2009c). Improved vocabulary production after naming therapy in aphasia: Can gains in picture naming generalise to connected speech? *International Journal of Language & Communication Disorders, 44*(6), 1036-1062.

Doyle, P. J., Goda, A. J., & Spencer, K. A. (1995). The communicative informativeness and efficiency of connected discourse by adults with aphasia under structured and conversational sampling conditions. *American Journal of Speech-Language Pathology, 4*(4), 130-134.

Edmonds, L. A., Nadeau, S. E., & Kiran, S. (2009). Effect of Verb Network Strengthening Treatment (VNeST) on lexical retrieval of content words in sentences in persons with aphasia. *Aphasiology, 23*(3), 402-424.

Edwards, S., & Tucker, K. (2006). Verb retrieval in fluent aphasia: A clinical study. *Aphasiology, 20*(7), 644-675.

Ferguson, A., & Armstrong, E. (1996). The PALPA: A valid investigation of language? *Aphasiology, 10*(2), 193-197.

Fillingham, J. K., Sage, K., & Lambon Ralph, M. A. (2006). The treatment of anomia using errorless learning. *Neuropsychological Rehabilitation, 16*(2), 129-154.

Fisher, C. A., Wilshire, C. E., & Ponsford, J. L. (2009). Word discrimination therapy: A new technique for the treatment of a phonologically based word-finding impairment. *Aphasiology, 23*(6), 676-693.

Franklin, S., Buerk, F., & Howard, D. (2002). Generalised improvement in speech production for a subject with reproduction conduction aphasia. *Aphasiology, 16*(10-11), 1087-1114.

Glosser, G., & Deser, T. (1991). Patterns of discourse production among neurological patients with fluent language disorders. *Brain and language, 40*(1), 67-88.

Haendiges, A. N., Berndt, R. S., & Mitchum, C. C. (1996). Assessing the elements contributing to a" mapping" deficit: a targeted treatment study. *Brain and language, 52*(1), 276-302.

Herbert, R., Best, W., Hickin, J., Howard, D., & Osborne, F. (2003). Combining lexical and interactional approaches to therapy for word finding deficits in aphasia. *Aphasiology, 17*(12), 1163-1186.

Herbert, R., Hickin, J., Howard, D., Osborne, F., & Best, W. (2008). Do picture-naming tests provide a valid assessment of lexical retrieval in conversation in aphasia? *Aphasiology, 22*(2), 184-203.

Hickin, J., Best, W., Herbert, R., Howard, D., & Osborne, F. (2002). Phonological therapy for word-finding difficulties: A re-evaluation. *Aphasiology, 16*(10-11), 981-999.

Hinckley, J., & Carr, T. (2005). Comparing the outcomes of intensive and non-intensive context-based aphasia treatment. *Aphasiology, 19*(10-11), 965-974.

Holland, A. L., Frattali, C., & Fromm, D. (1999). *Communication Activities of Daily Living: CADL-2*. Texas: Pro-Ed.

Howard, D. (2000). Cognitive neuropsychology and aphasia therapy: The case of word retrieval. In I. Papathanasiou (Ed.), *Acquired neurogenic communication disorders: A clinical perspective* (pp. 76-99).

Howard, D., & Harding, D. (1998). Self-cueing of word retrieval by a woman with aphasia: Why a letter board works. *Aphasiology, 12*(4-5), 399-420.

Jones, E. V. (1986). Building the foundations for sentence production in a non-fluent aphasic. *International Journal of Language & Communication Disorders, 21*(1), 63-82.

Kiran, S., Sandberg, C., Gray, T., Ascenso, E., & Kester, E. (2013). Rehabilitation in Bilingual Aphasia: Evidence for Within-and Between-Language Generalization. *American Journal of Speech-Language Pathology, 22*(2), S298-S309.

Leonard, C., Rochon, E., & Laird, L. (2008). Treating naming impairments in aphasia: Findings from a phonological components analysis treatment. *Aphasiology, 22*(9), 923-947.

Lind, M., Kristoffersen, K. E., Moen, I., & Simonsen, H. G. (2009). Semi-spontaneous oral text production: Measurements in clinical practice. *Clinical Linguistics & Phonetics, 23*(12), 872-886.

Linebarger, M. C., McCall, D., & Berndt, R. S. (2004). The role of processing support in the remediation of aphasic language production disorders. *Cognitive Neuropsychology, 21*(2-4), 267-282.

Links, P., Hurkmans, J., & Bastiaanse, R. (2010). Training verb and sentence production in agrammatic Broca's aphasia. *Aphasiology, 24*(11), 1303-1325.

Marshall, J. (1995). The mapping hypothesis and aphasia therapy. *Aphasiology, 9*(6), 517-539.

Marshall, J., Pring, T., & Chiat, S. (1998). Verb retrieval and sentence production in aphasia. *Brain and language, 63*(2), 159-183.

Mayer, J., & Murray, L. (2003). Functional measures of naming in aphasia: Word retrieval in confrontation naming versus connected speech. *Aphasiology, 17*(5), 481-497.

Milman, L., Vega-Mendoza, M., & Clendenen, D. (2014). Integrated Training for Aphasia: An Application of Part–Whole Learning to Treat Lexical Retrieval, Sentence Production, and Discourse-Level Communications in Three Cases of Nonfluent Aphasia. *American Journal of Speech-Language Pathology, 23*(2), 105-119.

Mitchum, C. C., & Berndt, R. S. (1994). Verb retrieval and sentence construction: Effects of targeted intervention. In M. J. Riddoch & G. W. Humphreys (Eds.), *Cognitive Neuropsychology and Cognitive Rehabilitation.* East Sussex: Lawrence Erlbaum Associates.

Mitchum, C. C., Greenwald, M. L., & Berndt, R. S. (2000). Cognitive treatments of sentence processing disorders: What have we learned? *Neuropsychological Rehabilitation, 10*(3), 311-336.

Murray, L., Timberlake, A., & Eberle, R. (2007). Treatment of underlying forms in a discourse context. *Aphasiology, 21*(2), 139-163.

Nicholas, L. E., & Brookshire, R. H. (1993). A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research, 36*(2), 338-350.

Nickels, L. (1992). The autocue? Self-generated phonemic cues in the treatment of a disorder of reading and naming. *Cognitive Neuropsychology, 9*(2), 155-182.

Nickels, L. (2002). Therapy for naming disorders: Revisiting, revising, and reviewing. *Aphasiology, 16*(10-11), 935-979.

Peach, R. K., & Reuter, K. A. (2010). A discourse-based approach to semantic feature analysis for the treatment of aphasic word retrieval failures. *Aphasiology, 24*(9), 971-990.

Prins, R., & Bastiaanse, R. (2004). Review: Analysing the spontaneous speech of aphasic speakers. *Aphasiology, 18*(12), 1075-1091.

Raymer, A. M., Ciampitti, M., Holliway, B., Singletary, F., Blonder, L. X., Ketterson, T., . . . Gonzalez Rothi, L. J. (2007). Semantic-phonologic treatment for noun and verb retrieval impairments in aphasia. *Neuropsychological Rehabilitation, 17*(2), 244-270.

Raymer, A. M., & Ellsworth, T. A. (2002). Response to contrasting verb retrieval treatments: A case study. *Aphasiology, 16*(10-11), 1031-1045.

Raymer, A. M., McHose, B., Smith, K. G., Iman, L., Ambrose, A., & Casselton, C. (2012). Contrasting effects of errorless naming treatment and gestural facilitation for word retrieval in aphasia. *Neuropsychological Rehabilitation, 22*(2), 235-266.

Raymer, A. M., Singletary, F., Rodriguez, A., Ciampitti, M., Heilman, K. M., & Rothi, L. J. G. (2006). Effects of gesture+ verbal treatment for noun and verb retrieval in aphasia. *Journal of the International Neuropsychological Society, 12*(06), 867-882.

Renvall, K., Nickels, L., & Davidson, B. (2013a). Functionally relevant items in the treatment of aphasia (part I): Challenges for current practice. *Aphasiology, 27*(6), 636-650.

Renvall, K., Nickels, L., & Davidson, B. (2013b). Functionally relevant items in the treatment of aphasia (part II): Further perspectives and specific tools. *Aphasiology, 27*(6), 651-677.

Robson, J. O., Marshall, J., Pring, T. I. M., & Chiat, S. (1998). Phonological naming therapy in jargon aphasia: Positive but paradoxical effects. *Journal of the International Neuropsychological Society, 4*(06), 675-686.

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language, 57*(3), 348-379.

Rose, M., & Douglas, J. (2008). Treating a semantic word production deficit in aphasia with verbal and gesture methods. *Aphasiology, 22*(1), 20-41.

Rose, M., Douglas, J., & Matyas, T. (2002). The comparative effectiveness of gesture and verbal treatments for a specific phonologic naming impairment. *Aphasiology, 16*(10-11), 1001-1030.

Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology, 22*(7-8), 691-706.

Schneider, S., & Thompson, C. (2003). Verb production in agrammatic aphasia: The influence of semantic class and argument structure properties on generalisation. *Aphasiology, 17*(3), 213-241.

Spencer, K. A., Doyle, P. J., McNeil, M. R., Wambaugh, J. L., Park, G., & Carroll, B. (2000). Examining the facilitative effects of rhyme in a patient with output lexicon damage. *Aphasiology, 14*(5-6), 567-584.

Thompson, C. (1989). Generalization research in aphasia. In T. Prescott (Ed.), *Clinical aphasiology* (Vol. 18, pp. 195-222). Boston: College Hill Press.

Thompson, C., & Shapiro, L. (2007). Complexity in treatment of syntactic deficits. *American Journal of Speech-Language Pathology, 16*(1), 30-42.

Thompson, C., Shapiro, L., Kiran, S., & Sobecks, J. (2003). The Role of Syntactic Complexity in Treatment of Sentence Deficits in Agrammatic AphasiaThe Complexity Account of Treatment Efficacy (CATE). *Journal of Speech, Language, and Hearing Research, 46*(3), 591-607.

Ulatowska, H. K., North, A. J., & Macaluso-Haynes, S. (1981). Production of narrative and procedural discourse in aphasia. *Brain and language, 13*(2), 345-371.

Waldron, H., Whitworth, A., & Howard, D. (2011). Comparing monitoring and production based approaches to the treatment of phonological assembly difficulties in aphasia. *Aphasiology, 25*(10), 1153-1173.

Wambaugh, J. L., & Ferguson, M. (2007). Application of semantic feature analysis to retrieval of action names in aphasia. *Journal of Rehabilitation Research and Development, 44*(3), 381.

Webster, J., Franklin, S., & Howard, D. (2007). An analysis of thematic and phrasal structure in people with aphasia: What more can we learn from the story of Cinderella? *Journal of neurolinguistics, 20*(5), 363-394.

Webster, J., Morris, J., & Franklin, S. (2005). Effects of therapy targeted at verb retrieval and the realisation of the predicate argument structure: A case study. *Aphasiology, 19*(8), 748-764.

Webster, J., & Whitworth, A. (2012). Treating verbs in aphasia: Exploring the impact of therapy at the single word and sentence levels. *International Journal of Language & Communication Disorders, 47*(6), 619-636.

Whitworth, A., Claessen, M., Leitão, S., & Webster, J. (under review). Beyond narrative: Is there an implicit structure to the way in which adults organise their discourse. *Clinical Linguistics & Phonetics*.

Whitworth, A., Leitão, S., Cartwright, J., Webster, J., Hankey, G., Zach, J., & Wolz, V. (this volume). Macrostructure as a scaffold to improving everyday discourse in aphasia: a multi-level approach. *Aphasiology*.

Youmans, G., Holland, A., Muñoz, M., & Bourgeois, M. (2005). Script training and automaticity in two individuals with aphasia. *Aphasiology, 19*(3-5), 435-450.

Table 1: A framework for describing generalisation and impact

GENERALISATION						IMPACT		
LEVEL	Word	Sentence	Connected Speech			Measures of everyday communication	Other report/ perception	Self-report/ perception
Elicitation methods	• Picture naming • Word association, e.g. noun to verb • Naming to definition • Sentence completion • Word fluency	• Constrained phrase or sentence production tests	Picture Description	Discourse		• Elicited production of everyday scenarios, e.g. ANELT, CADL-2, Scenario test • Rating scales e.g. CETI, FOQ-A	• Rating scales, e.g. Carer COAST, TOMS	• Rating scales, e.g. COAST, CDP, VASES • Quality of Life measures
				Monologue	Dialogue			
			• Complex picture description • Picture sequences	• Narrative, e.g. story retell • Personal narrative, e.g. recount • Procedural narrative • Expositions, e.g. opinions	• Conversation (more or less naturalistic sampling) • Role playing			
Focus	*Lexical change* • Treated & untreated words • Consideration of effects of semantic/ phonological similarity, typicality,	*Lexical change* • Treated & untreated words within sentences	*Lexical change* • Use of words e.g. number of words, type-token ratio, noun/verb diversity, • Words within sentence	*Lexical change* • Use of words e.g. number of words, type-token ratio, heavy: light verbs, noun: pronoun	*Lexical change* • Use of words e.g. number of words, type-token ratio, words/number of turns & substantive turns (see Herbert et al. 2008) • Words within sentence	Ability to convey message	Other's perception of change	Person's perception of change

	word class etc.	*Structural Change* • Sentence accuracy (thematic or syntactic)	*Structural change* • Structure of sentences e.g. MLU, proportion well-formed/ grammatical, • Sentence type e.g. verbs in correct argument structure, proportion of particular sentence type *Informativeness* • Relevant information e.g. correct information unit (CIU) • Ratings of understandability	*Structural change* • Macro structure e.g. cohesion, coherence, story grammar • Micro-structure e.g. structure of sentences, sentence type (see specific measures of structural change)	*Structural change* • Structure of sentences • Sentence type *Informativeness* • Understandability (see specific measures) *Change in interaction* • Conversation analysis			

Key to acronyms in table:

ANELT: Amsterdam-Nijmegen Everyday Language Test. Blomert, L., Kean, M.L. Koster, C. and Schokker, J. (1994) Amsterdam-Nijmegen Everyday Language Test: Construction, Reliability and Validity. *Aphasiology* 8, (4), 381-407.

CADL-2: Communication Activities of Daily Living-2. Holland, A. L., Frattali, C., & Fromm, D. (1999). Texas: Pro-Ed.

CDP: The Communication Disability Profile. Swinburn, K. and Byng, S. (2006). London: Connect.

CETI: Communicative Effectiveness Index. Lomas, J., Pickard, L., Bester, S., Elbard, H., Finlayson, A. and Zoghaib, C.. (1989) The Communicative Effectiveness Index - Development and Psychometric Evaluation of a Functional Communication Measure for Adult Aphasia. *Journal of Speech and Hearing Disorders* 54, no. 1: 113-124.

CIU: Correct Information Unit: Nicholas, L.E. and Brookshire, R.H. (1993). A System for Quantifying the Informativeness and Efficiency of the Connected Speech of Adults with Aphasia." *Journal of Speech & Hearing Research* 36, 338-350.

COAST: Communication Outcome After Stroke Scale. Long, A., Hesketh, A., Paszek, G., Booth, M., and Bowen, A. (2008) Development of a Reliable, Self-Report Outcome Measure for Pragmatic Trials of Communication Therapy Following Stroke: The Communication Outcome after Stroke (Coast) Scale." *Clinical Rehabilitation* 22,: 1083-1094.

Carer COAST: Long, A., Hesketh, A., & Bowen, A. (2009). Communication outcome after stroke: a new measure of the carer's perspective. *Clinical rehabilitation*, 23, 846-856.

FOQ-A: Functional Outcome Questionnaire for Aphasia: Glueckauf, R. L., Blonder, L. X., Ecklund-Johnson, E., Maher, L., Crosson, B., & Gonzalez-Rothi, L. (2003). Functional Outcome Questionnaire for Aphasia: Overview and preliminary psychometric evaluation. *NeuroRehabilitation, 18*(4), 281-290.

MLU: Mean Length of Utterance

Scenario Test: van der Meulen, I., van de Sandt-Koenderman W. M., Duivenvoorden, H. J. and Ribbers, G. M. (2010) Measuring Verbal and Non-Verbal Communication in

Aphasia: Reliability, Validity, and Sensitivity to Change of the Scenario Test. *International Journal of Language and Communication Disorders*, 45, (4) 424-35.

TOMS: Therapy Outcome Measures. Enderby, P. (1997) *Therapy Outcome Measures: Speech-Language Pathology*. London: Singular Publishing.

VASES: *Visual Analogue Self-Esteem Scale*. Brumfitt, S. and Sheeran, P (2010). Bicester: Winslow Press.

Table 2: Summary of effectiveness and within level generalisation in case series studies of single word noun and/or verb retrieval

Study	No. of Participants	Type of Therapy	Effect on Naming of Treated Words	Effect on Naming of Untreated Words
Hickin et al. (2002)	8	Phonological and orthographic cues for noun retrieval	No consideration of group performance 7/8 participants overall improvement 5/7 improved significantly more on treated nouns i.e. showed treatment effect	No consideration of group performance 1/8 significant improvement for untreated nouns
Abel et al. (2005)	10	Decreasing and increasing cues for nouns	Group analysis showed overall improvement and significant training effect. 8/10 participants showed overall improvement 5/10 showed significant training effect	Group analysis showed significant improvement in untreated items. 3/10 showed overall improvement on untreated items

				1/10 showed significant training effect on untreated items
Fillingham et al. (2006)	11	Errorless and errorful learning for noun retrieval	No consideration of group performance 9/11 showed significant gains for treated nouns	No consideration of group performance Minimal generalisation although gains for untreated nouns did reach significance at particular time periods for 5 participants
Raymer et al. (2006)	9	Gesture and verbal treatment for nouns and verbs	As a group, average effect size for treated nouns and verbs. 6/9 showed improvement (gains of >20% and effect size of >2.0) for treated words (either nouns or verbs)	As a group, no gains for untreated nouns or verbs. No improvement in retrieval for untreated words although 3/9 showed increase use of gesture for untrained verbs

Raymer et al. (2007)	8	Semantic phonologic treatment for nouns and verbs	No consideration of group performance 5/8 showed significant gains for treated nouns and verbs	No consideration of group performance No significant gains in retrieval for untreated nouns or verbs
Leonard et al. (2008)	10	Phonological component analysis for noun retrieval	No consideration of group performance 7/10 showed significant & large treatment effect for treated words	No consideration of group performance 3/7 showed significant gains on an unrelated naming test
Conroy et al. (2009a)	9	Errorful and errorless therapy for nouns and verbs	Group analysis showed significant effect of training with main effect of word class (nouns>verbs) with borderline effect of therapy type.	No consideration of group performance for untreated items

			All 9 participants showed significant gains for treated words	1/9 showed significant gains for untreated words
Conroy et al. (2009b)	7	Decreasing and increasing cue therapy for nouns and verbs	Group analysis showed significant effect of training with main effect of word class (nouns>verbs) with no effect of therapy type. All 7 participants showed significant gains for treated words	No consideration of group performance for untreated items No participant showed significant gains for untreated words
Raymer et al. (2012)	8	Errorless naming and gestural facilitation for nouns	No consideration of group performance for untreated items 6/8 showed improvements in treated words from at least one of the therapies (small to large treatment effects)	No consideration of group performance for untreated items 3/8 showed generalised naming improvements (small effect size)

Carragher et al. (2013)	9	Multicomponent verb therapy consisting of semantic feature analysis, gesture and phonemic cueing	Group analysis showed significant improvement in naming of treated items 8/9 showed significant gains for treated verbs	Group analysis showed significant improvement for untreated items that was maintained 1 month post-therapy 4/9 showed significant gains for untreated words immediately post-therapy but not maintained